DOCUMENT RESUME

ED 110 486 TM 004 759

AUTHOR Loveridge, Robert L.; And Others
TITLE The Evolution of the Role of a Formative Evaluator in an R & D Setting.
PUB DATE [Apr 75]
NOTE 17p.; Paper presented at the Annual Meeting of the American Educational Research Association (Washington, D.C., March 30-April 3, 1975)

EDRS PRICE MF-$0.76 HC-$1.58 PLUS POSTAGE
DESCRIPTORS Bias; Educational Development; Educational Research; Evaluation Methods; *Formative Evaluation; *Professional Personnel; *Program Development; Program Evaluation; *Staff Role

ABSTRACT

The role of the evaluator in a Research & Development (R & D) setting, particularly as seen in relationship to that of the product developer, is often viewed as adversal, rather than complemental. A case study approach has been employed in an attempt to define how the role as seen in the latter sense can best be implemented. Both failures and successes in R & D procedures provide inputs to the formulation of a well defined and viable evaluation role. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

Session # 25.13

ED110486

The Evolution of the Role of a Formative Evaluator in an R & D Setting

By

Robert L. Loveridge
Portland Public Schools

Sharon L. Rose
Philadelphia Public Schools

Marti Holmes
Research for Better Schools

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM 004 759

A Paper, presented at the 1975 Annual Meeting of the American Education Research Association, Washington, D.C.
March 30 to April 3, 1975.

Introduction

Every person involved in any field of evaluation evolves. Each person undergoes changes in the process of his career. The process of change begins when one decides that evaluation is going to be his or her career. For most this means entering an academic institution. During this time, each evaluator is born. His heredity and environment depend primarily upon the institution(s) he attends. The evaluator is taught and nurtured until he can walk and talk. However, the set of experiences he is exposed to are limited and controlled, and he enters the next era of change, the "world of practicing reality" without the opportunity for practice. In this world, changes begin to take place rapidly, and many of them can be drastic. One can only hope the "teachings of youth" are not totally forgotten and set aside. Each evaluator is in a different stage of development; some will progress faster than others, and some will either remain infants or die as functioning evaluators. For each evaluator, the vision or dream of a "better" future lies ahead, and to a great extent the future can be no better than dreams and visions that have been held for it in the past.

The Past

Upon arriving fresh from the academic world one brings many utopian ideas with the hope of improving the educational field. No task seems too large or too impossible; it should only be a simple matter of applying one of the many theories, models or designs that have been learned, and putting it to task. It is thought that the job of being a formative evaluator will be exciting, a position that will allow free flowing exchange, and/or modification for improvement. The evaluator thinks he is a co-captain or 1st officer of the ship, the instigator for change, the person providing information and recommendations for a "better way." All of this with the support of the product staff. Indeed, the evaluator sees himself or herself as a valuable asset to good product development.

Unfortunately, the world of textbooks and classrooms, often has little connection to the "nitty gritty" of the real world. Application often goes sour, and principles sometime fade and become compromised as the game of give and take begins.

Insights into the future of the evaluator can emerge when one returns to that eager young person about to venture into the "real world" of R & D evaluation. Except for a few obscure missing elements, his concept of evaluation has been stated very nicely by Scriven (1967). That is, evaluation involves the collection of information concerning the impact of an educational program. And, while there are many possible uses for such information, it is assumed by the new evaluator that the fundamental purpose of all evaluation is to produce information which can be used in educational decision-making. These decisions may be concerned with the continuation, termination, or modification of an existing program, or with the development and possible adoption of some new program. This definition appears to be sufficient until one discovers that those "few obscure missing elements," for the most part, can be translated into "the product developer." Normally before one is hired, the impression he had of his job was to: 1) familiarize himself with the product being developed, 2) test it in the field, 3) collect the data, 4) analyze the data, and 5) finally present what has been discovered to the product developer or employer for decision-making purposes.

When the evaluator commences his or her new job, the first task is to become familiar with the product or program to be evaluated. One assumes he or she is the specialist in this area (at least that was the impression conveyed at the time of hiring). After this, one begins the process of selecting one of the experimental designs to test the product. The design chosen has all of the best characteristics needed for a "good" evaluation. If needed, the design will include randomization of subjects to treatment, an "N" large enough to test for any treatment interaction(s), a balanced design, consultants to help with special problems, the same covariant(s)

for each subject, and costs, too, will be kept under control. This evaluation will be done by the books. The right way. And the evaluator will remain independent and totally unbiased. Unfortunately, the process isn't quite that simple.

The evaluator is to be the unbiased judge -- the essence of impartiality. However, this product has been nurtured from its inception by the developer. He is not unlike a protective parent who knows that his child must go to the hospital for a delicate and risky operation, but is uncertain of the possible outcome. The developer, too, knows that the product must be evaluated, but the outcome of that evaluation on his "creative effort" is as dreaded to him as the possibility of the death of the child on the operating table is to the parent. With these facts in front of the evaluator, he or she knows that a head-on-collision is not unlikely; the evaluator prepares for the collision with all the diplomacy and tact that is needed by a United Nations ambassador. The theories and studies previously read no longer represent reality, and the evaluator is suddenly face-to-face with a problem that did not exist back in the classroom. The present is here, and it must be dealt with now.

The Present

Initially, the role of the evaluator is ambiguous, for two masters, Research and Development, must be served. If the needs and demands of the two are congruent, the evaluator can proceed efficiently and effectively. But frequently the requirements of each are mutually exclusive. The researcher is seeking evaluation, and the developer may primarily be seeking confirmation.

The basic confrontation is between the theoretical framework which results from the evaluator's academic training and the constraints, demands, and idiosyncracies of reality. It is the traditional battle between theory and practice.

The academician can devise axioms and procedures for conducting research, including statements having to do with the random selection of subjects, the administration of evaluation instruments, and the exact method for implementing the

experimental treatment. But the practitioner must try to implement the theoretical requirements of the academician with the parameters of a real-life setting. The practitioner must deal with schools which refuse to permit random selection of students, or whose definition of randomness is significantly different from what is required for true randomness. He or she must deal with school strikes, early closings, with the replacement of a teacher in mid-treatment, with materials delayed or lost in transit, with unexpected amounts of attrition, with students who were contaminated by previous participation in a similar evaluation project, and with the way costs can spiral and consequently cause original plans to be modified in midstream. In fact, the possible violations to a tight research design are legion and it is experience which can best help the evaluator to prepare for them. An important caveat is that one cannot predict everything that will occur which may violate the design. However, one can attempt to control for as many as possible, remembering that these violations are not the ones learned about in the classroom.

There is probably no evaluator who has not been faced with these and even worse problems in the course of conducting an evaluation project. In fact, while these problems seem momentous to the new evaluator, they may appear trivial to experienced evaluators. Perhaps they are trivial in the face of higher order conflicts between the evaluator as evaluator and the evaluator as employee of the developer. The evaluator normally represents a certain threat to the developer. While the former may feel the need for conscientiously isolating variables, facts, and conclusions, the latter may interpret this as being negative. The unfortunate result of this perception is that the developer may internalize this negativism and see it as an attack on the program or product, and therefore, by extension, as a personal attack. This then, is one of the basic conflicts.

The friend vs. foe conflict that exists between the developer and the evaluator, when it exists, may provide the fuel for another basic conflict which has a tremendous impact on the findings of any evaluation. This is the introduction of bias.

While the evaluator is trained to be an objective observer, many developers have been emotionally involved in their programs. Thus, a seemingly innocent discussion of the merits of a program can become filled with hidden agenda. The new evaluator wants to prove his worth by finding out how successful the product is, as well as identifying areas within the product or program which need to be corrected in the course of its development; but the developer wants to protect his "baby" from attack. Far too often this results in a type of attack and defense behavior. The developer may attack the evaluator on an instrument or the design to be used in the evaluation. The evaluator naturally defends his approach, and in turn may attack the developer on some other point (e.g., the lack or unclarity of the objectives or outcomes to be measured by the evaluation). This process of attack and defense causes undue bias. The objectivity of the evaluator as well as the developer is put in jeopardy, and the evaluator, who may have been willing to modify his method, is now confronted with two issues that are hard to handle. One is that of his worth and integrity as a person or a professional, and the second is economic survival.

Obviously, these are generalizations, but beneath the objective facade of a developer insisting, "Of course I want to find out where and how I can improve this program," there lurks an artist, a creator, who doesn't want his creation to be stillborn. The goal of the evaluator, of course, is not to try to deliberately show the imperfections in the product, per se, but to display the true nature of the product and help improve it in its developmental stages by documenting its performance.

If the developer sees the evaluation as making a contribution to the overall improvement of the program, if his long-range objectives are seen as being congruent with those of the evaluator, if the results of the study are seen as being interim observations and not labels or immutable conclusions, then the basic conflicts can more readily be resolved. However, often, such is not the case.

Admitting that the above intentions and purposes are true is quite different from the actual application and operation of these intents and purposes in the planning and decision-making process.

The authors have synthesized their experiences with those related to them by other evaluators and have listed a number of typical kinds of situations in which evaluators often find themselves.

1) No objectives: The specific objectives of the developers are not made clear so the evaluator must "guess" at what specific elements are to be evaluated, or the objectives and goals are under continual revision and definition. Therefore, the emphasis of the evaluation is based on the evaluator's interpretation of the problem.

2) No communication: The developer may not be familiar with research terminology or procedures and may be reluctant to ask about them. The evaluator may assume the developer is aware of these things and may act accordingly, thus compounding the problem.

3) Super-objectivity vs. none: The evaluator may alienate the developer by seeming to be distant, unconcerned, and cold-blooded instead of an objective observer.

4) Poor timing: The evaluator is brought into the picture too late for any changes that may be necessary, and he is unable to evaluate the product on its merits or demerits for improvement. Indeed, he may really be asked to demonstrate only one or the other.

5) Emphasis on negativism: If the evaluator's role is seen as helping improve a program under development, the elements of the program which appear to be successful are often ignored and only those which show limited success are stressed. This emphasis makes the developer feel threatened and defensive about the program.

6) Exclusivity of evaluators or developers: Because evaluators and developers may not be able to communicate successfully with each other, they tend to discuss their work with only other evaluators or developers. This reinforces the feelings of "apartness" and reduces the opportunity for improving communication with each other. Evaluators and developers learn from other evaluators and developers, but they need to share their knowledge with each other.

7) Protection of feelings and protection of job: Because the evaluator has often been hired by or assigned to a developer there may exist the feelings that one should not be the bearer of unpleasant or even of non-positive tidings for fear of offending another professional or for fear of being fired for not producing supportive information.

When these are combined with a more exhaustive list prepared and discussed by Sanders and Guba (1973), the complexity of the job of the evaluator is indeed made manifest. The number of conflicts that must be faced seems almost endless, but they must be confronted and mastered. After determining the nature of the conflict(s) between the roles of evaluators and developers, one must plan how to eliminate any existing situations and how to prevent such situations from beginning in the future.

The Future

The past expectations and the present demands of the evaluator have probably been somewhat, but not entirely, exaggerated in this discussion. While it is true that every new evaluator as he leaves the world of academia may not have all of these expectations, it is likewise certain that every evaluator leaves with some if not many of them. Similarly, all the problems alluded to in the "present" are not faced by each evaluator on each project, but that one or more of these "problems" or "conflicts" or an adaptation of them, will at some time have to be dealt with is almost a certainty.

How each person reacts to the frustrations created when he or she meets with the conflicts of expectancy or theory, and with the daily demands of "practical

reality" (i.e., how the system is really being operated) will, in the long run, determine the future of the evaluator as a functioning member of the development team. If one feels that nothing can be done except accept the "functioning reality" of product evaluation, then his or her contribution to "real" evaluation is, at best, limited and many times totally fruitless. On the contrary, if one sets the theory or expectation of "good" evaluation high, then in spite of any consequences that may follow, "better" evaluation will be the most probable outcome.

Achieving "better" evaluation to a large extent merely means the controlling of biases that are present or that can be introduced into any evaluation which is to be conducted. Bias, much like Bermuda grass, is rooted in more than one place, and it is hard to control. However, one way to locate a large set of roots contributing to evaluation bias is to look at the organizational model under which the evaluator must function (see also Scriven, 1974).

For the purpose of this paper a Section Developer is the person developing or supervising the creation or development of an educational product (e.g., new curricula, strategy, process, etc.). A Program Director is one who directs two or more Section Developers; a Division Director is in charge of two or more Programs; and the Corporation consists of two or more divisions or programs, depending on the size of the Corporation. Solid lines represent both supervision and budgetary responsibility. Dotted lines represent information and/or indirect supervision responsibilities, but not budgetary control. (Refer to Models 1 through 5).

Scriven (1974) has stated two general principles that must be kept in mind if bias is to be minimized. The first is the "Principle of Independent Feedback," and the second is the "Principle of Independence Maintenance." The first principle requires that "no unit in the chart [operational hierarchy] should rely entirely for evaluative feedback about a given sub-unit on the pipe down which it pours the money for that same sub-unit" (Scriven, 1974). The second principle requires

evaluation of the evaluation because without it, independence, whenever it exists, tends to be temporary in any bureaucratic structure.

It should be obvious that as the model number increases the likelihood of controlling bias in any evaluation also increases. It is not until you reach the second model that you even obtain the first principle, and Model 3 is the first time for the second principle. It is not until Model 5 that both principles are met for the corporation, and then the first principle is violated for the funding agency. However, it should be easier for the funding agency to have greater "confidence" in the results, findings, and decisions reached in Model 5 than in any other model presented.

Since it is unlikely in any event that bias will be controlled or eliminated merely by the adoption or implementation of any of the models described, it becomes necessary to look for other means to achieve "better" evaluation results and controlling biases.

The following represent possible suggestions for implementation to achieve "better" evaluation: The first and most obvious, is the need to have better training programs (Worthen, 1972 and Scriven, 1974). Not only must the evaluator receive better training in the what, why, and how of evaluation, but this training must include more doing of evaluation. Each evaluator must not only be taught how to walk and talk, but he must be given a chance to practice the new found "powers" or "skills." In this way he or she is more likely to know what is going to trip him and how to move, circumvent, or eradicate the object that tripped him. New evaluators, as well as seasoned ones, "fail to produce satisfactory evaluations because they have not been equally well tutored in the practical problems that they are likely to encounter under real world conditions" (Sanders & Guba, 1973). Indeed, many of these problems transcend most technical aspects of any evaluation to include the introduction of more general or pervasive variables. Here is the place for academic institutions and institutions involved in R & D to cooperate.

The R & D institutions need to make a commitment, and to this extent so must the funding agencies, to assist in the training of new evaluators. The academic institutions must in turn require practical experience as part of any degree program. This idea is not a new one; it has been practiced for years in the training of doctors, pharmacists, teachers, and skilled craftsmen of all trades.

A second method of improving evaluation is the creation of evaluation guidelines. Worthen (1972) and Scriven (1974) are just two of many who advocate this position. Good guidelines would aid evaluation and help control bias in at least two ways: 1) the evaluator would have a base from which he can develop a comprehensive evaluation plan; 2) it would establish with the developer the requirement that the evaluator must perform certain activities and that these activities are done to achieve more usable results from which to base decisions.

Third, evaluation must be included from the beginning of the development project (Worthen, 1972). This would help alleviate many of the problems encountered by the evaluator (e.g., unmeasurable objectives or goals, arriving on the scene too late, etc.).

Fourth, every evaluator should recognize that no matter how hard he or she tries to be objective it is practically impossible to eliminate all the biases he or she has or that are some way forced upon him or her. Consequently, every possible effort should be made to supplement his own evaluation efforts with some type of outside evaluation. These might include one or more of the following as resources permit: 1) Goal-Free Evaluation (GFE) as proposed by Scriven (the articles referenced in the back of this paper provide a good explanation of GFE, its purposes and objectives, and dated 1972, 1973, & 1974); The Adversary Hearing (a good discussion of this process is in the article by Owen, 1973); or contracting of part of the evaluation outside the parent unit.

A fifth way of controlling bias (mainly the bias of how the evaluation data are used in decision-making for product modification) is to bring in another developer

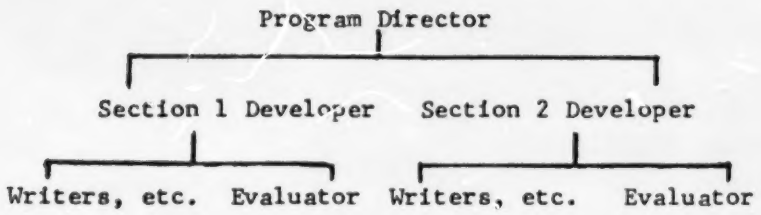
in the field. This developer would review the evaluation data and make recommendations in light of that data. This may be hard for some to accept, but it is asking no more of the developer than the developer asks of the evaluator when he calls in a consultant to review the evaluator. If you wish, this might be termed Developer-Free Evaluation.

Last, and most important, the evaluator and the developer should not look at each other as members of opposing teams who are trying to score the most points. They are really both members of the same team, holding different positions, but whose end purpose is the same. One should not be considered superior to the other, but that each has an important and essential function to perform if a good educational product is to be produced.

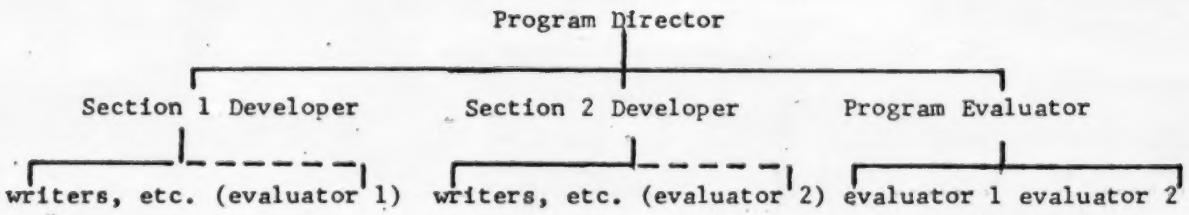
Summary

Evaluation performs a unique function in product development as well as in establishing product worth and effectiveness. The elimination of bias is particularly hard in most organizations where product development is conducted, but its control is to be strived for at all levels and by each concerned party. The establishment of safeguards coupled with a conscientious effort to control biases will aid both in the better utilization of evaluation findings and the prevention of their misuse. Research, Evaluation, and Development personnel should not be considered as different or opposing forces in product development, but as a team striving for creative, efficient, effective, and worthwhile educational products and programs.

Model 1



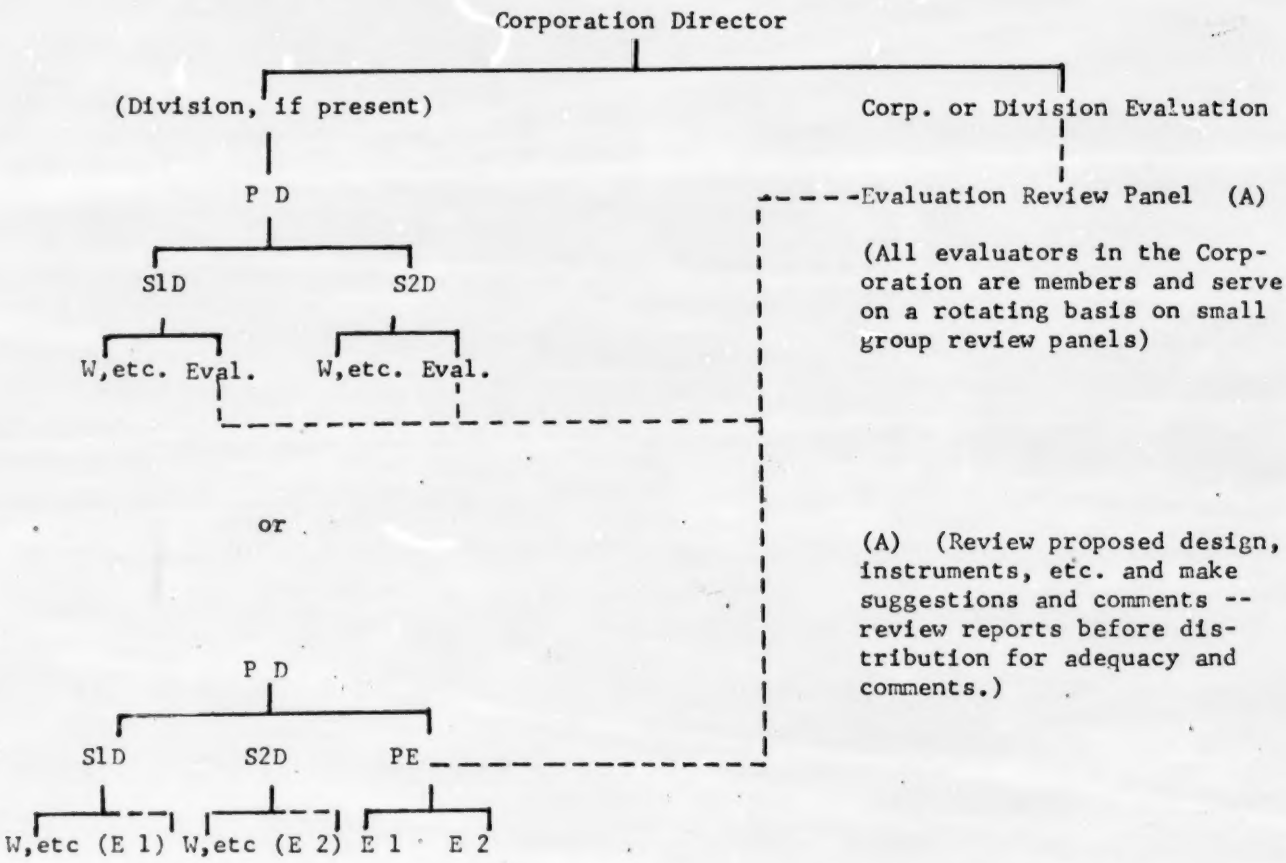
Model 2



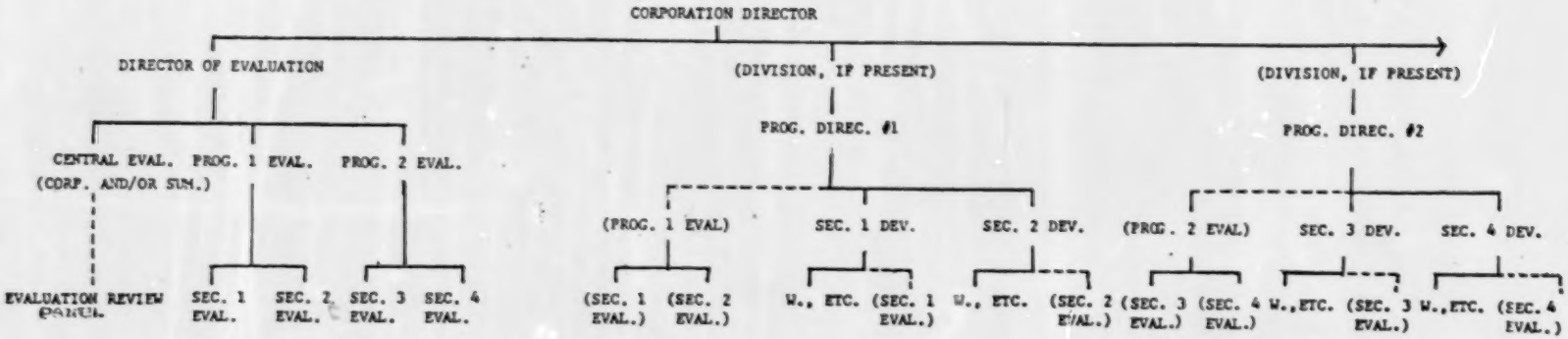
Model 3

This model could be the same as Models 1 or 2, with the added dimension of Division or Corporate evaluation of the evaluator on a periodic basis.

Model 4


Corporation Director
(Division, if present)
Corp. or Division Evaluation
P D
Evaluation Review Panel (A)
(All evaluators in the Corporation are members and serve on a rotating basis on small group review panels)
S1D
S2D
W,etc. Eval.
W,etc. Eval.
or
(A) (Review proposed design, instruments, etc. and make suggestions and comments -- review reports before distribution for adequacy and comments.)
P D
S1D
S2D
PE
W,etc (E 1)
W,etc (E 2)
E 1
E 2

MODEL 5


CORPORATION DIRECTOR
DIRECTOR OF EVALUATION
(DIVISION, IF PRESENT)
(DIVISION, IF PRESENT)
CENTRAL EVAL. (CORP. AND/OR SUM.)
PROG. 1 EVAL.
PROG. 2 EVAL.
PROG. DIREC. #1
PROG. DIREC. #2
EVALUATION REVIEW PANEL
SEC. 1 EVAL.
SEC. 2 EVAL.
SEC. 3 EVAL.
SEC. 4 EVAL.
(PROG. 1 EVAL)
SEC. 1 DEV.
SEC. 2 DEV.
(PROG. 2 EVAL)
SEC. 3 DEV.
SEC. 4 DEV.
(SEC. 1 EVAL.)
(SEC. 2 EVAL.)
W., ETC.
(SEC. 1 EVAL.)
W., ETC.
(SEC. 2 EVAL.)
(SEC. 3 EVAL.)
(SEC. 4 EVAL.)
W., ETC.
(SEC. 3 EVAL.)
W., ETC.
(SEC. 4 EVAL.)

REFERENCES

Owens, T. R., "Educational Evaluation by Adversary Proceeding." School Evaluation: The Politics and Process. House, ed. McCutchan. 1973, pp. 285-305.

Sanders, J. A. & Guba, E. C. "A Taxonomy of Problems Confronting the Practitioner of Education Evaluation." Sept. 1973.

Scriven, M. "Conversational Contact - Technical." Evaluation. Vol. 2, No. 1, pp. 9-16, 1974.

Scriven, M. "Evaluation Bias and Its Control." University of California at Berkeley. February, 1974.

Scriven, M. "Methodology of Evaluation." American Education Research Monograph Series on Curriculum Evaluation. Chicago: Rand McNally, 1967.

Scriven, M. "Prose & Cons About Goal-Free Evaluation." The Journal of Educational Evaluation. Vol. 3, No. 4, Dec, 1972.

Worthen, B. R. "Impediments to the Practice of Educational Evaluation." Laboratory of Educational Research, Research Paper No. 60, University of Colorado. April 1972.

RLL:ms